ABSTRACT
dissertation for the degree of Doctor of Philosophy (PhD)
in the educational program 8D06101 - "Information systems (by industry)"

ISMUKHAMEDOVA AIGERIM MELSATOVNA

ALGORITHMIC SUPPORT OF AN INTELLIGENT CLINICAL DECISION SUPPORT SYSTEM

Relevance of the research topic. With the development of information technology, the medical industry is undergoing a significant transformation due to the digitalization of large data sets. These changes offer new opportunities to improve the quality of medical care, but also pose challenges related to the need to effectively analyze and interpret complex and diverse data. In particular, the rapid increase in the amount of available medical information requires the development of new approaches and data processing methods that can handle these challenges at a qualitatively new level.

This study fits into the large-scale project on digitalization of the medical industry in the Republic of Kazakhstan (Kazakhstan 2050, eHealth, Digital Kazakhstan, Electronic Health Passport, implementation of the Smart City concept) and concerns the implementation of a clinical decision support system based on artificial intelligence within the national healthcare system. Such introduction of innovative technologies in medicine will not only improve the quality of patient care through earlier and more accurate diagnosis, but also contribute to the optimization of administrative processes, guaranteeing reliable and efficient management of medical data.

Digital transformation of healthcare is the process of introducing modern digital technologies and innovations into the healthcare sector in order to improve the quality of healthcare services, improve access to them, optimize the management of medical processes and increase the efficiency of healthcare institutions. This process involves the use of new technologies to improve diagnosis, treatment, monitoring, and data and patient management. One of the key aspects of digital transformation of healthcare is the use of machine learning algorithms in medical data analysis, such as image interpretation, disease diagnosis, disease prognosis, and treatment optimization. Machine learning algorithms can help in processing large amounts of data, which speeds up the decision-making process.

The application of artificial intelligence and machine learning in processing data from patients with chronic ischemia and other medical fields has significantly improved outcomes in diagnosis, therapy, and disease prognosis. Julian Mutz and Catherine Lewis used machine learning techniques to determine biological age from data on mental characteristics associated with accelerated aging. Artificial intelligence and machine learning are playing a key role in developing novel approaches to treat Parkinson's disease and early diagnosis of high-mortality cancers.

The theory and practice of introduction of information technologies in the medical sphere, regarding diagnosis and adjustment of treatment of various diseases are considered in the works of Kazakh scientists M.T. Ipalakov, M.E. Mansurov, A.K. Mukashev, N.P. Saparkhodjaev, A.K. Mukashev and others.

Studies show that most machine learning models achieve an accuracy of more than 95% in the diagnosis of chronic lymphocytic leukemia, and also provide 100% accuracy in differentiating this disease from other pathologies. Machine learning and deep learning methods are actively used to close diagnostic gaps in cases of hereditary arrhythmias. These technologies also play an important role in selecting optimal algorithms for breast cancer treatment.

To improve results, researchers often use combinations of 5-6 machine learning models, including basic classification methods within deep learning. Indian experts are applying a range of machine learning classifiers including NB, KNN, SVM and decision tree algorithms such as ID3 and C4.5 to predict and diagnose diabetes. Colleagues from Pakistan demonstrate the application of six well-known machine learning algorithms such as SVM, KNN, logistic regression, decision trees, random forest and naive Bayes for diabetes prediction, achieving accuracy up to 77%.

In this study, the field of Diabetology was selected to investigate clinical decision making processes. Diabetes mellitus, commonly known as diabetes, is a serious global health problem affecting millions of people worldwide. The disease has reached epidemic levels in many regions of the world, with its prevalence continuing to rise. According to international medical data from the World Health Organization, approximately 422 million people globally suffer from this disease, which is approximately 6.028% of the total population. In the context of the Republic of Kazakhstan, WHO statistics show that 11.5% of the population suffers from diabetes mellitus; among them, 11.7% are women and 11.3% are men. The National Register of the Republic of Kazakhstan for 2021 records 317,597 registered cases of diabetes mellitus, including 314,407 adults, 2,379 children under 14 years of age and about 811 adolescents in the age category of 15 to 17 years.

Swedish scientists are investigating the application of machine learning in preventive programs for type 2 diabetes by identifying key risk factors for the development of this disease.

Thus, the relevance of the topic is that machine learning and artificial intelligence are becoming key technologies in the development of techniques and approaches in medical decision support systems. Government programs serve as a platform for implementing and testing these advanced technologies, and the more extensive the database of patient data, the higher the accuracy of analysis and the faster the implementation of new unique software products and solutions in the field of disease diagnosis and prognosis.

The relevance of the dissertation work is also confirmed by the fact that the study was carried out within the framework of Scientific Research Work under the contract No. 321/23-25 dated 03.08.2023 on the topic AR19679525 "Program complex of diagnostics of clinical and hematological syndromes for electronic health passport", carried out within the framework of the budget program "Grant

financing of scientific research", as well as Scientific Research Work under the contract No. 128/JG 5-24-26 dated 02. 06.2024 on the theme AR22683316 "Application of machine learning algorithms for medical decision support systems" carried out within the framework of the budget program "Grant financing of young scientists under the project 'Zhas galym' for 2023-2025 years".

The object of the study is the clinical decision support system.

The subject of the study is the algorithmic support of intelligent clinical decision support system in endocrinology and diabetology.

The idea of the work is the application of machine learning and artificial intelligence technologies for the tasks of intelligent support of clinical decision-making processes in endocrinology and diabetology, allowing to contribute to the implementation of the global strategy of the World Health Organization on e-health for 2020-2025.

The aim of the study is to develop algorithms for intelligent clinical decision support based on machine learning algorithms.

In order to achieve the goal, the following research and main tasks should be performed :

- study of diagnostic processes in the information-clinical field and features of clinical decision support systems;

- study of the issues of efficiency of use and problems of application of machine learning technology for clinical decision support systems;

- development of a conceptual model of the clinical decision support process based on EDA methodology;

- development of a hybrid algorithm for clinical decision support based on under sampling and automatic optimization of parameters;

- development of an algorithm for applying the Grid Search method for clinical decision support tasks based on the CNN model;

- development of an algorithm for ensembling neural network architectures for clinical decision support tasks;

- development of the information model of the intellectual system of clinical decision support;

- implementation of EDA methodology on clinical data of endocrinology and diabetology;

- conducting an experimental study of the endocrinology clinical decision support algorithm based on endersampling technology;

- evaluating the effectiveness of the diabetes prediction algorithm based on the deep neural network model with optimized hyperparameters;

- evaluation of the accuracy of implementation of the algorithm for ensembling LSTM and RNN neural network architectures for clinical decision support tasks;

- developing the architecture of a clinical decision support system.

Scientific novelty of the dissertation research consists in the fact that for the first time to improve the efficiency of clinical decision support processes in endocrinology and diabetology a set of algorithms integrating the technology of andersampling and ensembling of LSTM and RNN neural network architectures is proposed.

The main scientific statements put forward for defense:

- *a hybrid algorithm for* clinical decision support based on undersampling and automatic parameter opitimization;

- *algorithm of* Grid Search method application for clinical decision support tasks based on CNN model;

- *algorithm of* ensembling neural network architectures for clinical decision support tasks.

The practical significance of the results of the dissertation research lies in the applicability of the proposed set of algorithms integrating the technology of andersampling and ensembling of LSTM and RNN neural network architectures for solving problems of information and analytical support of managerial decision-making in the support of technological business processes in endocrinology and diabetology.

The scientifically substantiated theoretical and experimental results of the thesis work were used in thescientific project on the topic "Software complex of diagnostics of clinical and hematological syndromes for electronic health passport", carried out within the framework of the budget program "Grant financing of scientific research", as evidenced by the certificate of participation in the project.

It should be noted that the experimental results of the work were used in the scientific project on the theme "Application of machine learning algorithms for medical decision support systems", carried out within the framework of the competition for grant financing of young scientists' research under the project "Zhas galym" for 2024-2026, as evidenced by the certificate of participation in the project.

Certificates of state registration of rights to copyright objects №49449 from 04.09.2024 "Algorithm of support of clinical decisions on the basis of technology of andersampling", as well as №4737 from 01.08.2019 "Program module of diagnostics of clinico-hematological syndromes" and №41784 from 05.01.2024 "Database of differential diagnostics of clinico-hematological syndromes on the basis of algorithm of morphological classification" are received.

The modules developed in the thesis work on methods of big data management, including their collection, storage and processing, on the basis of work with medical data in the context of CDSS was successfully applied on the basis of production of "ЮвентаМед" LLP, as evidenced by the act of introduction.

The methods of IT project management and ethical aspects of data use developed in the dissertation work were implemented in the educational process of education program 7M04104 "IT Management", KAFU for 2024-2025 years in the form of disciplines "Design and Implementation of Software System" and "Digital business modeling", as evidenced by the act of implementation.

The results of the thesis research were implemented in the educational process in 2023-2024 academic year in the following courses of lectures and

practical classes education program "Mathematical and computer modeling", D. Serikbayev East Kazakhstan technical university, namely in the disciplines "Modeling of biological processes" and "Basics of neural networks", as evidenced by the act of implementation.

Research Methods. The work uses methods of machine and deep learning, methods of statistical analysis, methods of big data processing, decision-making theory, as well as methods of experimental research and modeling of the system architecture.

Approbation of the results of the dissertation research. The main results of the dissertation work were reported at scientific seminars of the department "Information Systems" D. Serikbayev East Kazakhstan technical university and at the following international scientific-practical conferences: "ADVANCED SCIENCE" (Penza, Russia, 2017); "4th International Conference on Computer and Technology Applications" (Istanbul, Turkey, 2018); "4th International Conference on Engineering and MIS" (Istanbul, Turkey, 2018.); "IV International Scientific and Technical Conference of Students, Graduates and Young Scientists" (Ust-Kamenogorsk, Kazakhstan, 2018); "XVIII International Research Competition 'Best Scientific Paper 2018' (Penza, Russia, 2018); "Computational and Information Technologies in Science, Engineering and Education: 9th International Conference, CITech 2018" (Ust-Kamenogorsk, Kazakhstan, 2018); "Application of Information and Communication Technologies-AICT 2018" (Almaty, Kazakhstan, 2018); "Application of Information and Communication Technologies-AICT 2018" (Almaty, Kazakhstan, 2018); "5th International Conference on Engineering and MIS" (Nur-Sultan, Kazakhstan, 2019); "12th IEEE International Conference 'Application of Information and Communication Technologies - AICT2019' (Baku, Azerbaijan, 2019); 'VI International Scientific and Technical Conference of students, graduate students and young scientists 'Creativity of the young - innovative development of Kazakhstan'" (Ust-Kamenogorsk, Kazakhstan, 2020); "2021 International Young Engineers Forum (YEF-ECE)" (2021); "7th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)" (Ankara, Turkey, 2023).

Author's personal contribution. Problem statement, formalization of all considered problems, search for methods and algorithms of their solution, as well as scientific and practical results given in the thesis, their analysis, formation of final conclusions were carried out personally by the author of the thesis.

Publications on the topic of the dissertation research. On the subject of the dissertation published 28 scientific papers, including 8 in scientific journals recommended by the Committee for Control in the field of science and higher education MES RK; 12 in the proceedings of international conferences; 3 publications indexed in the database SCOPUS; 1 monograph.

Publications published within the framework of scientific research in scientific editions Scopus and Web of Science were cited 49 times.

The results of the experimental study of algorithms of intellectual support of clinical decision-making systems were described in the article "Integrating machine learning in electronic health passport based on WHO study and healthcare

resources", published in the journal "Informatics in Medicine Unlocked", which has a CiteScore of 86 in Scopus in the direction of "Computer Science".

Structure and volume of the dissertation work. The dissertation consists of an introduction, three chapters, conclusion, bibliography and appendices.

The first section investigated the existing approaches to diagnosis, identifying the key challenges and benefits of using clinical decision support systems.

The second section focused on clinical decision support algorithms and models. The developed conceptual model and algorithms, such as the Undersampling method and Grid Search technique for neural networks, demonstrated their practical relevance and feasibility of integration into existing clinical systems. The ensemble of different neural network architectures has shown its ability to improve the accuracy of diagnostic predictions, which is key to improving the quality of health care. The experimental study conducted in section three confirmed the effectiveness of the proposed algorithms on clinical data. The accuracy evaluation of the LSTM and RNN models showed that these techniques can significantly improve diagnostic performance and clinical decision support.

The conclusion of the thesis work shows the main results, draws conclusions on the thesis research, thus confirming the scientific novelty and practical significance of the study.